

# Beyond Words of Popularization Mining: Reviews on Comic Books Movies

Angela S. H. Lee\*, Kaza NS. M. Kamil and Shankaraar Narendranaath

Department of Computing and Information Systems, School of Science and Technology,  
Sunway University, Subang Jaya, Malaysia; angelal@sunway.edu.my,  
kazanizamshah@gmail.com, shannaren10@gmail.com

## Abstract

**Objectives:** This paper aims to analyse the terms used by the of movie critics on popular comic book movies. From the analysis we are able to know what are the factors that most movie critics will focus at, whether the top movie critics review has an impact on the decision of general movie goers and from this will determine why the movies was a box office success. **Methods:** The research is done via text mining from SAS Enterprise Miner software which utilize the text analytics tool to analyse movie reviews by movie critics' professionals. The analytical processes involved are text parsing, text filtering, text clustering and the text topic method. Correlation of terms is identified to help in determining the significance of each terms nested within the reviews. **Findings:** The findings show that the actors, the movie's characters and the film storyline are the most common type of terms that most movie critics are focusing on, thus this factors is what drives the reputation of the film and making it a blockbuster hit around the world. **Application/Improvements:** The results obtained can help the entertainment industry on their decision making on what to focus on when it comes to producing comic books superheroes into movies based on the sentiment analysis in identifying the postive and negative terms and their relationship with one another.

**Keywords:** Actors, Characters, Comic Book, Movie Reviews, Superhero, Text Mining

## 1. Introduction

With the advancement of technology for example Web 2.0 has given user the interaction they need, since it is all about participation of the users. Websites like Amazon or Internet Movie Database (IMDB) is allowing users to post their reviews. This reviews will be shared among other users and readers, in which is gives them the knowledge and insight of the topic being reviewed. In the 21<sup>st</sup> century there is a lot of Electronic Commerce websites, potential users will always read reviews from other users that have purchased the same products they are interested in and they can either look at the review from the website or into social network, blog or even discussion forum.

When compared to product reviews, movie reviews involve, for example, the plot, visual effects, settings, soundtrack and editing. Also it involves people related to the movie like directors, screenwriter and most importantly the actors and actresses. Overall, movie review is very much

complex and it is richer in content; thus it might be a challenge to analyze movie review compared to product review. This research will focus on three type of movies which are based on comic movies released from February to April of 2016. The paper is organized as follow: Section 2 Literature Review, Section 3 Research Methodology, Section 4 Analysis and Results, Section 5 Discussion and Section 6 Conclusions.

## 2. Literature Survey

### 2.1 Knowledge Model

Knowledge modelling objective is to give a description of organizational knowledge in which the knowledge is being processes by the computers from their topics of interest are visualize. In other words, the knowledge modelling is used to capture and model a selected knowledge into a format that is reusable for the reason of sharing or even preserving.

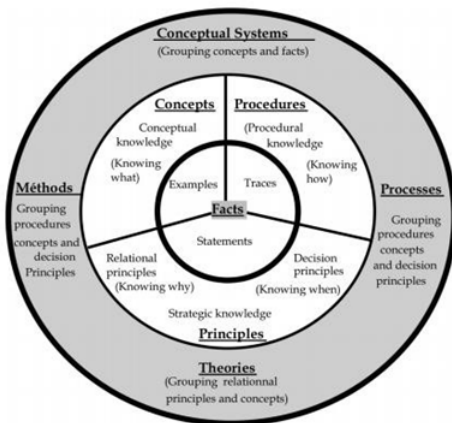
\*Author for correspondence

When it comes to the knowledge management, the crucial part is the knowledge modelling<sup>1</sup>. It is argued that the knowledge models give the organization the better understanding of the overall working mechanisms that surround the entire system of the knowledge-based for example knowledge domain, the schemas, methods and also tasks<sup>2</sup>.

### 2.1.1 Taxonomy of Knowledge

Figure 2 shows the four composite systems which are theories, processes, methods and conceptual systems. All four types of the composite systems around found in every areas of knowledge to which it can give meanings or definitions within the knowledge classification. The conceptual systems are a large number of concepts that are connected with each other. Methods and Processes involve the connection and grouping the procedures concepts and decision principles. Theories refers to grouping or the relational principles and concepts<sup>3</sup>.

What is taxonomy? It is considered the Knowledge Organization System (KOS)<sup>4</sup>. Taxonomy is a set of words which has been organized and sorted in order to check the use of terms that is being used in a particular subject field into a vocabulary so that it is able to go through the process of storing and retrieving of items from a certain repository. Knowledge domain, subject area, enterprise area or even a topical are considered KOS. With taxonomy it is a vocabulary that is being controlled in such a way that is focuses on concepts. It is not specific items but it caters to the concepts which are being represented by words and terms. Taxonomies allow information to be more accessible. Information which is well organized into a logical, consistent framework has such high value for the purpose of retrieving that information and even doing analysis on them.



**Figure 1.** Taxonomy of knowledge: Facts and abstract knowledge.

## 2.2 Text Mining and Sentiment Analysis on Movie Reviews

Research study done by<sup>5</sup> where it looks into movie reviews. With the use of SAS Enterprise Miner, they are able to generate summaries of terms along with the relationship that surrounds the aspects of movies. Movie reviews are able to give the audience the understanding of what they like or even disliked in a particular movie. Furthermore, by using text mining, they were able to find terms that commonly used in the reviews and how does this reviews eventually affect the reputation of the movies. They are able to determine the satisfaction or even dissatisfaction of the customers in relationship to the movie and this may affect the revenue that is generated by the movie itself whether it is in a positive way or negative way. Besides using text mining, the researchers also use sentiment analysis where models on existing review can be build where it predicts the new reviews is bad or good. Overall with sentiment analysis it will help in revealing why a particular movie it a hit or a flop. This will also help movie makers in improving the quality of the movie in order to meet the audience expectation.

In this research, three (3) specific comic book movies that were released in the beginning of 2016 were analyzed. With text mining it will help in determining specific terms that critic has said in their review regarding the comic book movie, thus helping in answer why are comic books movies getting so much hype and popularity. Basically we are seeing the connection of “Deadpool”, “Batman v. Superman: Dawn of Justice” and “Captain America: Civil War”, the relationship of the 3 major comic book movies that was released this year. When it comes to business intelligence, text mining plays a very large role for any kind of organizations as it enables them to do a thorough analysis towards their customers or even competitor<sup>6-9</sup>. Text mining is considered a powerful tool as it provides insights that are valuable to the business for example, if looking at the customer satisfaction they are able to find ways to improve it and from there gaining the competitive advantage.

## 2.3 Information Extraction and Values of the Data Extracted

When it comes to analysing unstructured text the initial step is information extraction. The main part of this step is acknowledge and validate phrases and terms which had meaningful relationships between them<sup>12</sup>.

The information gathered is automatically stored in a database constructed by the information extracting module in a structured form<sup>12,16</sup>. Clear-cut output and queries that are accurate are the major advantage of information extraction. Moreover, they can be reviewed efficiently and can be displayed visually on screen<sup>14</sup>. All of this extracted information helps the people to do their work faster as they are not required to read or go through all the documents one by one. Instead what they are receiving in the end is the essence of what holds within the text. All of this data are directly created by the customers or by any relevant stakeholders which in returns gives us an understanding of the reality that we could not have had in the past years<sup>5</sup>.

## 2.4 Comic Book Movies and their Impact on Mainstream Media

Comic books have proven their cultural legitimacy and impact on pop culture as we know it. It has been implied that comic-books have been at the forefront of human evolution in terms of cultural impact and visual media as a whole<sup>12</sup>. Over the years, this underlying thread of comic books has continued with more volume as well as visual styles from outrageous mythological images of Jack Kirby or the intimate autobiographical stream-of consciousness of Robert Crumb throughout the 1940s, 1960s, 1980s and the 2000s.

With this in mind, the power that comic books had, it was only logical that at some point, the natural progression of this medium was to the big screen. However, it does beg the question, how are comic book movies so widely popular in mainstream media? Why could it be dominating the box-office at the moment? According to an article by Nathalie DuPont, comic books and their superheroes encapsulate the typical box-office fare. In the 1950s, the beginning of the television in everyday life meant that Americans started spending more time at home<sup>13</sup>. As a result, film goers dropped significantly in the 1960s. However, the 1970s drew audiences back into the theatre following the introduction of the modern-day "blockbuster" from hits such as Jaws and Star Wars<sup>14</sup>. Hence analysis was made and common thread with these successful blockbuster films was that all of them had fairly simplistic characters, loud music, fast-paced action, many special effects and merchandising benefits<sup>15</sup>. When Richard Donner made the first Superman movie, it basically ticked all the boxes of this formula. The film had a fairly simplistic character, (Clark Kent/

Superman), loud music (iconic score by John Williams), fast-paced action (the character of Superman had super-speed and durability), many special-effects (Superman flying scenes) and merchandising benefits (Superman action figures). What are other factors that could be the reason? Maybe the adventures of comic book superheroes are made for the modern exhibition market? Throughout the 1980s and 1990s, along with the changing world of comic books, the number of multiplexes increased more audience involvement etc<sup>16</sup>.

According to the article, comic book movies are able to touch upon human sensibilities as well as cultural reference, for example, the Dark Knight and its mirroring of the Iraq War, Iron Man and the villains being middle-eastern. In the X-Men franchise, fellow mutants continue to fight each other, proving superiority of one group or the other, when really both sides of the coin are very similar to each other. Both factions are mutants who want fit in.

## 3. Research Methods

### 3.1 Text Mining Process

The sole purpose of Text Mining is to process data text information that is unstructured. Basically it is extracting text which is meaningful. Information can be imported or extracted from various sources to derive summaries for the entire words. So, generally it is used to determine the similarities between the documents that are being analysed.

#### 3.1.1 Text Import

The first step in the text mining process is to import the text. This research focus on crawling public reviews on 2 public forum website. Critic reviews are exported manually by copy paste in Excel sheet and converting the Excel into SAS executable file. A total of 1000 reviews have been gathered as can be seen in Figures 2 shows the Excel format being imported into SAS Enterprise Guide so that it can be saved as SAS file in order to mine using SAS Enterprise Miner.

#### 3.1.2 Text Parsing

The Text Parsing node as shown in Figures 6 and 7 allows to parse any documents for a really detail information regarding the terms. Basically text parsing decomposes the textual data. Furthermore, it helps in modifying the terms which has been parsed by allowing users to drop terms.



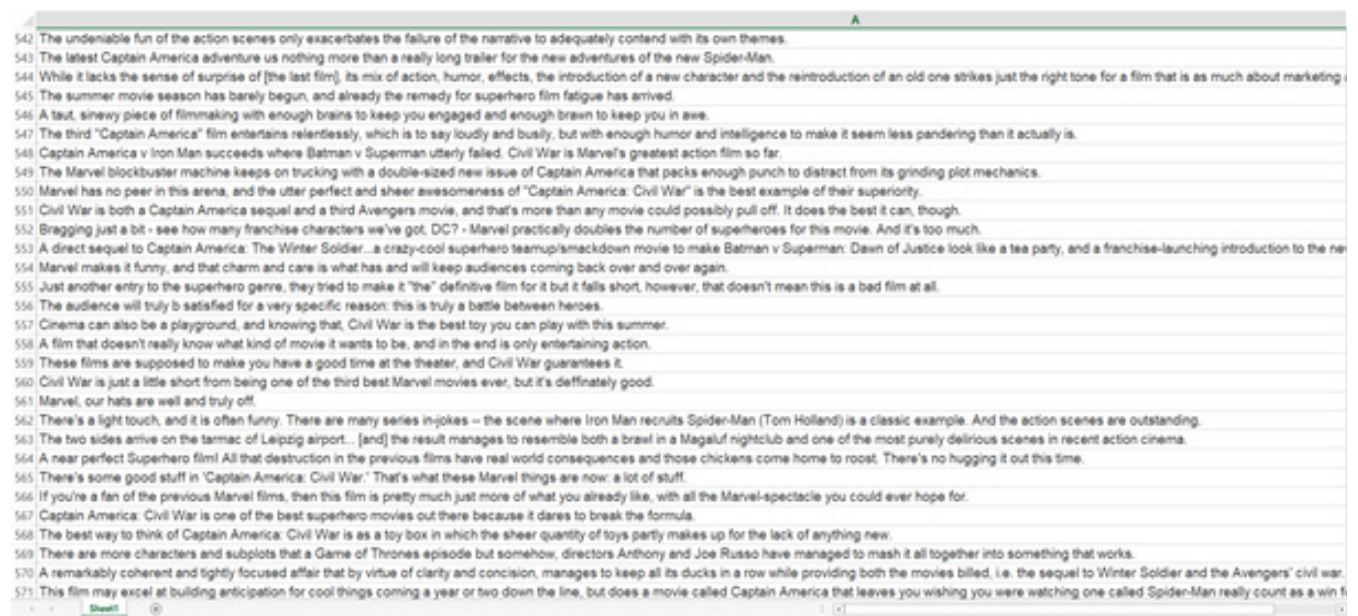


Figure 2. Excel format of critic reviews.

### 3.1.3 Text Filter

The use of the Text Filter is to reduce the total number of terms that has been parsed or analyzed documents. It helps in eliminating any information that users may find to be extraneous, thus only keeping the most valuable and relevant information. Besides that, it helps in removing unwanted terms.

### 3.1.4 Text Cluster

Text Cluster basically clusters the documents into fragments and only reports on descriptive terms for those clusters. Figure 8 shows the Text Cluster node.

### 3.1.5 Text Topic

When it comes to Text Topic, this stage helps user to explore the document where it automatically associate documents and even terms according to user defined topics.



Figure 3. Text parsing in text mining.



Figure 4. Text filter in text mining



Figure 5. Text cluster in text mining.

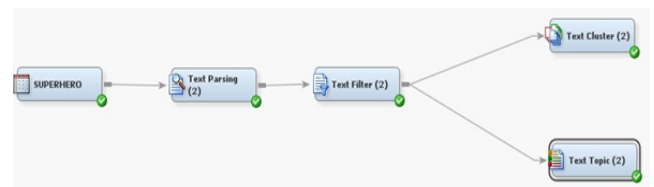


Figure 6. Text topic in text mining.

The main goal is to create list of topics in which users are able to establish combination of words that they are interested in analysing as shown in Figure 9.

## 4. Research Results and Analysis

### 4.1 Movie A

This analysis focuses on the first comic book based movie released during the year 2016, so based on the result shown in the Figure 7, the frequency of all the terms used in the critic reviews. From that the observation that can be made the highest frequency term is “be”. That is why the term “be” is not kept because it is not relevant to the entire movie and in the Text Parsing process the term and eventually drop the term since there is no significance.

The most linked terms to the term to this action comedy movie are “over-the-top”, “universe”, “antidote”, “character”, “superhero”, “strong”, “time” and “wink” are display in a concept map. All of the terms are all related to this movie. Next is drilling down into the second link of the term on the concept map. The first term is “over-the-top” as shown in Figure 8. So the drill down shows that the term “over-the-top” is linked to hoot and violence. In the Movie A it is undeniable that the violence is over-the-top. Hoot is referring the character of the main cast in the movie in which his sense of humour was really over-the-top. Hoot refers to somebody who is hilarious and has a great sense of humour. Therefore, the concept map shows

a relation between hoot and violence which indicates the movie shows violence and it is about a superhero that has a hoot characteristic.

Movie A despite being rated “restricted” because of the strong language and violence used throughout the movie it is only suitable for age 18 and above, however it has garnered positive reviews from critics praising the performance of the main actors in playing the character and this movie has hit blockbuster chart. Critics were really amused and fond of the character portrayed as he has a great sense of humour. Overall Movie A is superhero movie that is focused more towards romance, comedy, and action.

Term	Role	Attribute	WEIGHT	Freq	# Docs	Keep	Rank for Variable NUMDOCS	+movie,+superhero movie,+superhero,finally,deadpool	reynolds,ryan,performance,+role,+star	fun,irreverent,viole,t,+deliver,+film	+genre,+superhero genre,+superhero,subversive,+mean
deadpool ... Prop	Alpha		0.136532	144	142Y		1	0.111	-0.043	-0.006	0.039
+ movie ... Noun	Alpha		0.233812	93	84Y		2	0.543	0.061	-0.037	-0.025
+ superhero ... Noun	Alpha		0.244761	80	77Y		3	0.482	-0.025	-0.003	0.223
reynolds ... Prop	Alpha		0.336048	45	45Y		4	-0.036	0.625	0.02	0.001
+ film ... Noun	Alpha		0.343897	47	44Y		5	-0.037	-0.008	-0.07	0.015
fun ... Noun	Alpha		0.460665	29	24Y		6	0.02	-0.017	0.853	0.008
ryan ... Prop	Alpha		0.445689	24	24Y		6	0.074	0.564	-0.011	-0.022
+ character ... Noun	Alpha		0.453112	23	23Y		8	0.052	0.015	-0.013	-0.016
+ good ... Adj	Alpha		0.463625	23	22Y		9	-0.068	0.105	0.036	0.08
+ genre ... Noun	Alpha		0.477489	20	20Y		10	-0.001	0.03	0.022	0.661
marvel ... Prop	Alpha		0.482847	22	20Y		10	0.01	0.111	-0.059	-0.066

Figure 7. Highest Terms Frequency.

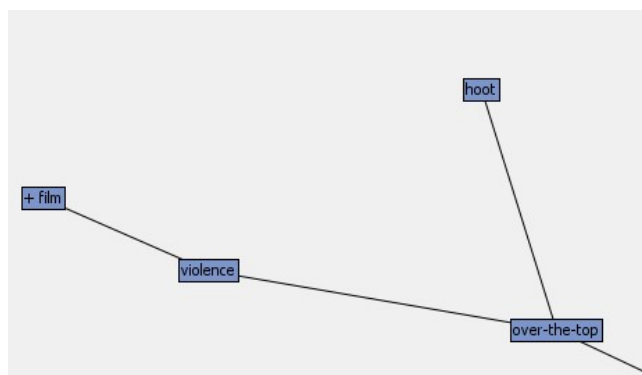


Figure 8. Concept linking term over-the-top.

## 4.2 Movie B

The reviews of this film has been mixed between critics and fans, with mostly negative reaction critically while being more evenly split in terms of the reaction from the fan base. These are the two most popular comic-book characters.

Figure 9, the “material” sub-node had the sub-nodes “good”, “character”, “action”, “deliver” and “enjoyable”.



Figure 9. Concept linking term material.

This could indicate that the sub-nodes of material were descriptors of the material being the film itself. As the film was mostly plagued with negative reviews, the “good” sub-node can be assumed that the majority of the reviews had the common belief that the “good” elements were the visual aspects of the film and the special effects. “deliver” sub-node indicate that most of the reviewers felt that the director for the most part fan expectation, it may have been ambitious and may be different from most comic book movies.

### 4.3 Movie C

As displayed in Figure 10, the frequency of all the terms used in the critic's reviews is sorted in descending order. It is observed that "be" has the highest frequency among all the words used, similar to the case of Movie A and B. The term "civil" has the fifth highest frequency, which has the frequency of 118 and a weightage of 0.183. The reason the term "civil" is at a lower frequency level is due to the fact that the critic reviews are about the Movie C, then the importance of the terms must be related to the movie and what critics are reviewing about. Based on this the highest weightage would be the term "good" because it is an adjective, and therefore is a good representation of the general atmosphere of the movie.

Term	Role	Attribute	Freq	# Docs	Keep	Parent/Child Status	Parent ID	Rank for Variable NUMDOCS
+ be	Verb	Alpha	198	148Y	+		54	1
civil	Adj	Alpha	118	118Y	+		10	2
is	Noun	Alpha	147	116N			12	3
war	Noun	Alpha	115	115Y	+		11	4
civil war	Noun Group	Alpha	113	113Y	+		2	5
+ movie	Noun	Alpha	100	89Y	+		152	6
america	Prop	Alpha	86	84Y	+		58	7
+ not	Adv	Alpha	69	63N	+		34	8
captain	Prop	Alpha	63	61Y	+		121	9
+ good	Adj	Alpha	59	52Y	+		505	10
+ action	Noun	Alpha	54	54Y	+		204	11
marvel	Prop	Alpha	53	52Y	+		110	12
+ superhero	Noun	Alpha	50	47Y	+		47	13
+ film	Noun	Alpha	49	46Y	+		27	14
+ do	Verb	Alpha	45	37N	+		372	15

Figure 10. Terms.

As seen from the drill-down of "winner", the terms "big", "deliver", "audience", "picture" and "emerge" are linked to "winner" as displayed in Figure 11. From this, an inference can be made that Movie C was a hit with critics as one can say that it emerged as a big hit and delivered to audience expectations. Therefore, it can be concluded that this was a successful movie.

Movie C was considered a big hit among critics, as it portrays to very strong characters. This movie already

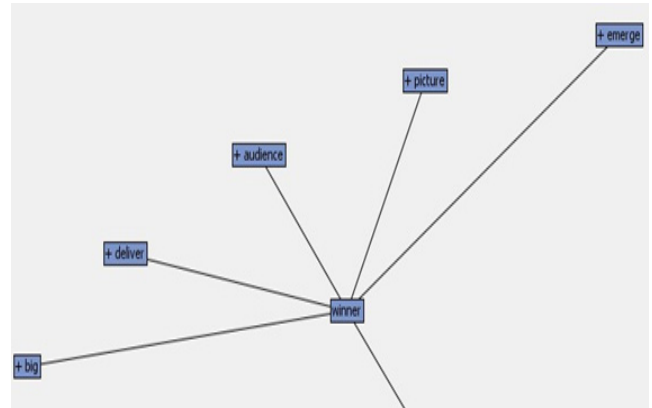


Figure 11. Concept linking term winner.

has its momentum due to its past sequels and stand-alone movies. This movie was very much anticipated by not just movie critics but also movie goers. Overall critics agree that an entire film delivered well in its storyline and the story arc of each character was good.

### 4.4 Analysis on 2016 Comic Book Movies

Figure 12 shows that the term "movie" has the highest frequency which is 266, the second highest is "superhero" and third is "film". This shows that the research is on the right track since this is looking into superhero movies or films.

Under the "interesting" sub-node were the sub-nodes "idea", "action", "story", "character" and "movie" as shown in Figure 13. From this, it can be deduced that what may have remained a common thread behind all these movies in terms of what was interesting was probably the idea behind the movie, the action, the story, the character that would affect the movie in general.

Term	Role	Attribute	WEIGHT	Freq	# Docs	Keep	Rank for Variable NUMDOCS	+character, +fan, title, +hero, avengers	civil, war, civil war, america, captain	justice, dawn, batman, superman, batman	+superhero movie, +superhero, +movie, +want, different
+ movie	Noun	Alpha	0.212817	266	239Y		1	0.112	-0.054	-0.012	0.484
+ superhero	Noun	Alpha	0.25823	176	170Y		2	-0.065	-0.024	0.027	0.522
+ film	Noun	Alpha	0.277902	169	153Y		3	-0.075	-0.056	0.025	-0.006
deadpool	Prop	Alpha	0.281194	146	144Y		4	0.018	-0.001	0.025	0.073
civil	Adj	Alpha	0.309273	118	118Y		5	0.028	0.5	0.009	0.009
war	Noun	Alpha	0.313001	115	115Y		6	0.028	0.5	0.01	0.001
civil war	Noun Group	Alpha	0.315541	113	113Y		7	0.032	0.495	0.01	0.013
+ good	Adj	Alpha	0.332503	109	103Y		8	0.008	0.031	0.048	-0.163
superman	Prop	Alpha	0.333089	106	102Y		9	0.102	0.008	0.372	-0.021
america	Prop	Alpha	0.358014	87	85Y		10	-0.078	0.272	0.002	0.044
batman	Prop	Alpha	0.360507	82	84Y		11	0.052	0.024	0.002	0.004

Figure 12. Terms for all films.



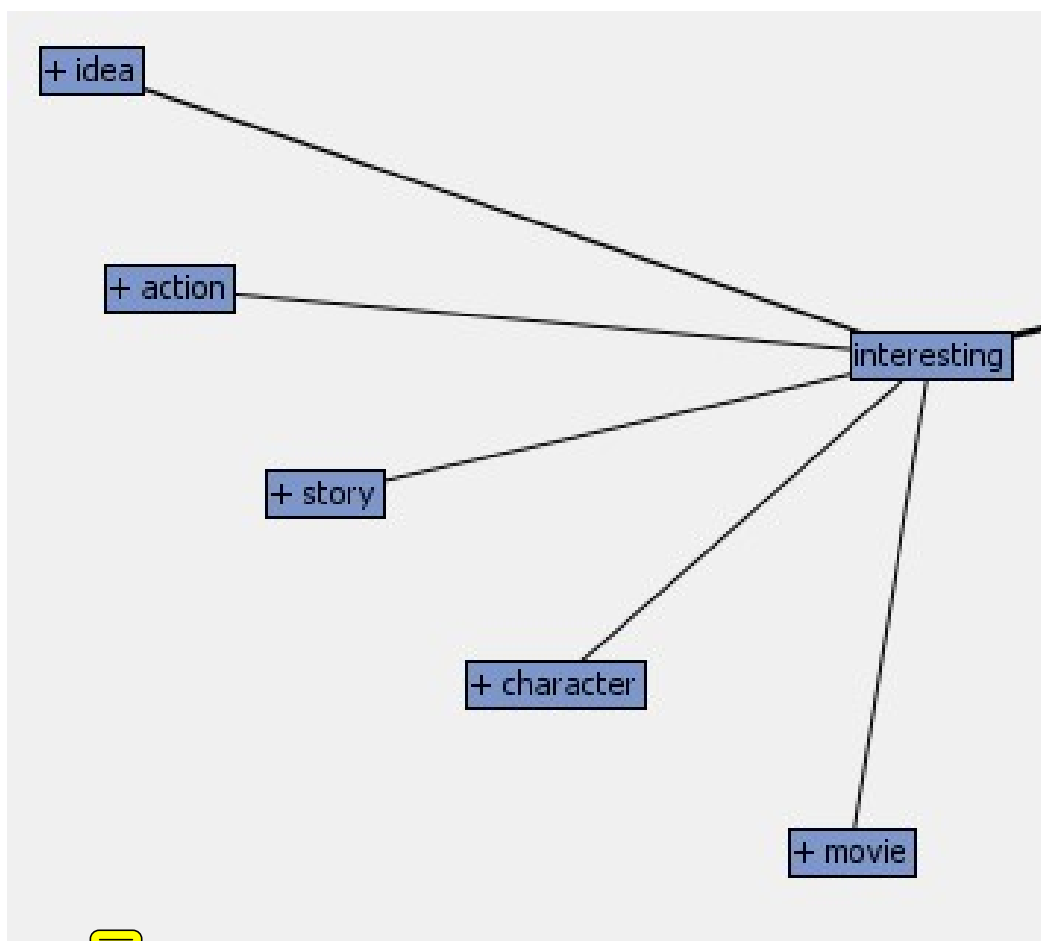


Figure 12. Concept linking term interesting.

Cluster ID	Descriptive Terms	Frequency	Percentage
1	film marvel universe brand +well marvel galaxy cinematic interesting +stop +theme series humor charm +idea far ...	91	9%
2	fun funny +lack +character +scene +question script +flight +filmmaker +exhaust +joke +entertain +action +hero cgi ...	280	28%
3	half hour +hour half two-and-a-half +two-and-a-half hour tedious +head truth little +fine cgi +effect +bore +confuse +...	21	2%
4	'superhero movie' +superhero deadpool +movie +book comic outrageous +comic book movie +appeal +sexy origin di...	114	11%
5	tim 'superhero genre' affleck +big miller +genre +marvel +director +audience snyder dc +credit +franchise +blockbuste...	197	20%
6	batman dawn justice superman league steel woman zack dc +look +dream +spectacle snyder +want +flaw ...	89	9%
7	role reynolds ryan +play mutant deadpool +wisecrack charm +performance +find +funny origin +dull +actor +credit ...	42	4%
8	'civil war' america anthony captain civil joe war russo iron +late +fine comic +director avengers +good ...	166	17%

Figure 13. Cluster ID.

Figure 13, the highest frequency of terms by cluster, would be cluster ID 2 (character), 5 (superhero genre) and 8 (good) with frequencies 280, 197 and 166 respectively.

In the text cluster, in the 3 highest frequencies of text clusters are the following terms;

Cluster ID "character"

- character
- scene
- script

Cluster ID "superhero genre"

- action
- genre
- audience

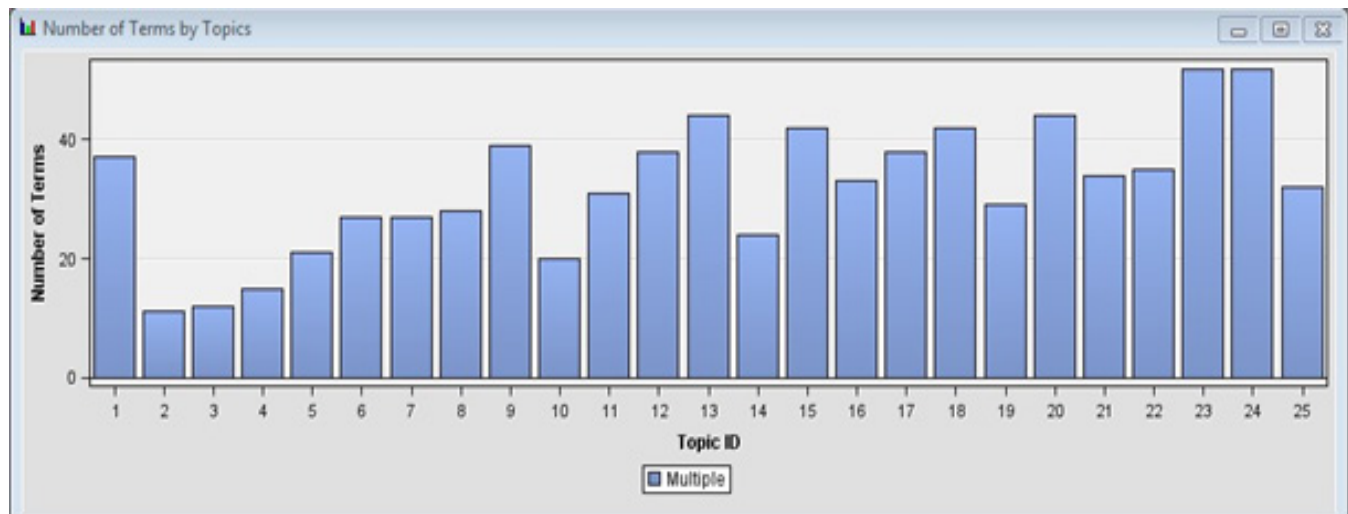


Figure 14. Number of terms by topic.

Topics

Category	Topic ID	Document Cutoff	Term Cutoff	Topic	Number of Terms	# Docs
Multiple	1	0.092	0.053+character,+fan,title,+...	37	94	
Multiple	2	0.155	0.050civil,war,civil war,amer...	11	117	
Multiple	3	0.124	0.049justice,dawn,batman,...	12	63	
Multiple	4	0.123	0.050+superhero movie,+s...	15	86	
Multiple	5	0.100	0.050snyder,zack,+director,...	21	56	
Multiple	6	0.102	0.050reynolds,ryan,+play,+r...	27	55	
Multiple	7	0.097	0.052marvel,universe,+goo...	27	85	
Multiple	8	0.100	0.051+book,comic,+comic...	28	40	
Multiple	9	0.092	0.053+action,humor,plot,+d...	39	99	
Multiple	10	0.093	0.050fun,funny,violent,unev...	20	61	
Multiple	11	0.090	0.050+hour,half,+half hour,...	31	32	
Multiple	12	0.099	0.052deadpool,funny,+funn...	38	142	
Multiple	13	0.088	0.054dc,marvel,universe,wo...	44	96	
Multiple	14	0.096	0.052+film,+deliver,plot,hug...	24	148	
Multiple	15	0.089	0.053+big,+audience,+deliv...	42	72	
Multiple	16	0.103	0.053batman,superman,+...	33	92	
Multiple	17	0.095	0.054america,captain,+thin...	38	108	
Multiple	18	0.102	0.053+good,+movie,marvel,...	42	107	
Multiple	19	0.096	0.052+genre,superhero ge...	29	59	
Multiple	20	0.095	0.053+scene,+great,+flight,+	44	76	
Multiple	21	0.087	0.053+story,origin,+feel,+id...	34	62	
Multiple	22	0.100	0.054+good,+time,deadpoo...	35	117	
Multiple	23	0.089	0.054+blockbuster,+fan,cap...	52	114	
Multiple	24	0.088	0.054+hero,+thing,funny,+bl...	52	54	
Multiple	25	0.090	0.051russo,joe,anthony,+dir...	32	41	

Figure 15. Topic ID.

Cluster ID “good”

- franchise
- blockbuster
- good

From this observation, it can be concluded that majority of critics mostly disregard the source material (comic book) in favour of the quality of the overall film.

The highest number terms by topic would belong to terms with topic ID 23 (blockbuster) and 24 (hero), with

52 terms each as can be seen in Figure 14, the Topic ID in detail based on the graph bar are displayed in Figure 15.

## 5. Discussion

From the findings, the different metrics of the terms mentioned can be used to evaluate all these movies and how they are related in terms of the review.

One observation that was made was that even though these are very different comic book movies from different

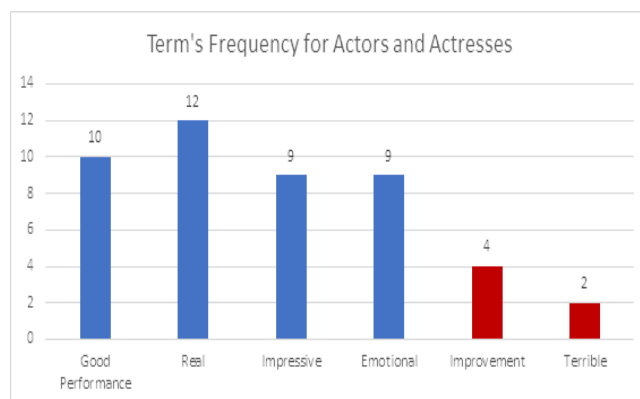


studios, the type of terms that remain common in at least 2 of all 3 films mentioned were;

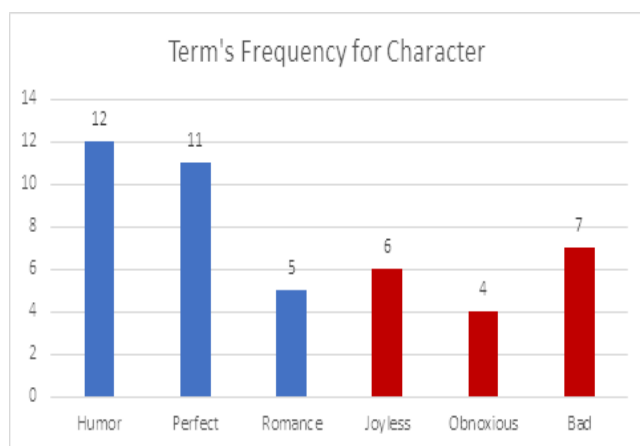
- The actors names ("Reynolds"/"Affleck").
- The characters ("character"/"America").
- The film ("movie"/"film").

This could mean that the critical reaction to these films was mainly on the characters, the actors playing the characters and how it works as a film.

Table I shows some of the representative reviews related towards the actors whether it is positive or negative. Some critics praise the actors/actress performance in portraying the characters while some critic on their poor acting performance.



**Figure 16.** Positive and Negative frequency terms related to "Actors and Actresses".



**Figure 17.** Positive and Negative frequency terms related to "Character".

Figure 17 shows the positive and negative terms related to the movie character. Most of the critics uses "Humor" as one of the positive terms to describe the movie character and "Bad" as one of the negative terms. Table II relates to the characters from the comic books portrayed within the movies. Most critics do like how the characters are portrayed and each main characters were justified enough throughout the film.

**Table I.** Examples related to actors/actresses

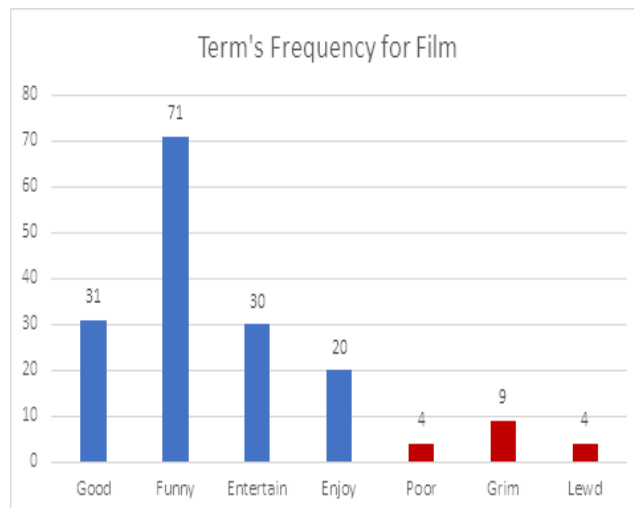
Topics	Examples
Movie critics sharing positive reviews	<i>Does justice to the character in this hilarious R rated superhero movie.</i>
	<i>Cool action, a strong performance worth checking out on the big screen.</i>
Movie critics sharing negative reviews	<i>Needs a new acting.</i>

**Table II.** Examples related to characters

Topics	Examples
Movie critics sharing positive reviews	<i>As overwhelming as all this seems, the story stays incredibly focused on what defines a superhero.</i>
	<i>It makes all the other superhero characters look like relics from the 1950s.</i>
	<i>Stark's frustration at the rigidity and short-sightedness of his confreres and his anguish at where it all leads are vivid and felt.</i>
Movie critics sharing negative reviews	<i>Hates his employer.</i>
	<i>A reference isn't a joke, and I didn't find the movie or the character - funny at all.</i>

Figure 18 shows that majority of the positive sentiment terms were represented by the term " Funny" and " Good".Whereas for negative sentiment terms was " Grim" and " Poor". This simply means majority movies critics uses the terms " Funny" + "Good" + "Entertain" + "Enjoy"

in representing an overall film critics. Table III looks into some of the reviews regarding the film as a whole. While some criticize the storyline of some of the movies, but mostly praised the comic books movies for their action and portrayal of the characters.



**Figure 18.** Positive and Negative frequency terms related to the overall "film".

**Table III.** Examples related to the film

Topics	Examples
Movie critics sharing positive reviews	<i>The film is a spectacular achievement, a true portrayal of the character's irreverence with own irreverence, a comedy that defies genres and an action movie that don't get lost in the cliché.</i>
	<i>A film so fun you will leave the theatre sayin: "lets go get some chimichangas!".</i>
	<i>I appreciated for not only some great action scenes, but also for delving into more internal issues with our heroes, allowing us to see them from an alternate angle than what we're used to..</i>
Movie critics sharing negative reviews	No amount of crash, bang or wallop can conceal this confused and ungainly storyline.
	<i>It is intermittently rewarding, but ultimately a bit disappointing.</i>

Perhaps critics view comic-book movies as character driven pieces and as in how well the actors played the character were emphasised. It also can be assumed that most critics do not necessarily care much about the source material, which why the terms "film" or "movie" comes up more often than not. It can be assumed that maybe critics view and rate comic-book movies in terms of how well it works as an overall movie, rather than whether or not it was adapted to an acceptable degree from the comic book sources.

This observation still holds water as when the text mining analysis was done for all the movies combined, in the concept map, under the "interesting" sub-node were the sub-nodes "idea", "action", "story", "character" and "movie".

All 3 of these movies was a box office success. Movie A had a budget of \$58 million<sup>1</sup> and during the opening weekend the total number of worldwide sales was \$264.7 million<sup>4</sup>. Movie B budget was \$250 million<sup>2</sup> and made a total of \$422.5 million during their opening weekend worldwide sales<sup>4</sup>. Movie C also had a budget of \$250 million<sup>3</sup>, in return their opening weekend worldwide sales was \$379.5 million<sup>4</sup>. The total box office sales for each movie at the end was \$783.1 million for Movie A, \$873.6 million for Movie B and \$1.153 billion for Movie C<sup>1-3</sup>.

From this, we can deduce that what may have remained a common thread behind all these movies in terms of what was interesting and also what made the movie has such a good reputation in the box office was probably;

- The idea behind the movie.
- The action.
- The story.
- The characters that would affect the movie in general.

## 6. Conclusion

According most of the reviews that were analyzed, it seems that more emphasis was placed on the characters when talking about the film. The prospect of seeing the audience owns beloved characters driving plot were a plus point in a movie.

Another observation that can be concluded from this study is the quality of the comic-book movies is based on the film itself and not the source material itself.

We can conclude that factors such as characters, actors and the quality of the film were a few factors that

may affect the popularity of comic book movies positively or negatively.

In the end, from the perspective of the movie critics, it can be said that all the factors mentioned above is what made all three of the movies a success in the box office as they had made millions and also billions. So movie goers end up going to the cinema to watch the movies due the characters, actors, and the overall quality of the film. Storyline much did not bother the movie goers from not spending their money in watching the movie even though movie critics criticize some of the movie's poor storyline. Top movie critic's review thus not affect the overall general movie goer's decision in watching the movie. From this text analytics analysis, we are able to identify the different terms and segment each terms to positive and negative categories. We learned that most of the movies critics' concern improved the production and story line of the movies.

## 7. References

1. Anon. *Deadpool* (2016) - Box Office Mojo. Boxofficemojo.com. 2016.
2. Anon. *Batman v Superman: Dawn of Justice* (2016) - Box Office Mojo. Boxofficemojo.com. 2016.
3. Anon. *Captain America: Civil War* (2016) - Box Office Mojo. Boxofficemojo.com. 2016.
4. Anon. All Time Worldwide Opening Records at the Box Office. Boxofficemojo.com. 2017
5. Bagga, S. Text Analytics: Unlocking the Value of Unstructured Data, International Institute For Analytics. 2016.
6. Dupont N. Hollywood adaptations of comic books in a post-9/11 context: The economic and cultural factors. 2016.
7. Gabilliet J, Beaty B, Nguyen N. *Of Comics and Men: A cultural history of American Comic Books*. 1st ed. MS: University Press of Mississippi. 2010; 1(1):85–8.
8. Jadhavar A, Sirolkar P, Chakraborty G. Analysis of IMDB reviews for movies and television series using SAS Enterprise Miner and SAS Sentiment Analysis Studio. 2016. p. 1–17.
9. Lee A, Lim T. Opinions from University Students' feedback using Text Analytics. *The Journal of Information Technology and Industry*. 2016; 4(1):26–33.
10. Lee A, Lim T, Leow S, Lim J. Second International Conference on Digital Information and Communication Technology and its Applications (DICTAP); 2017. p. 183–8.
11. Lee A, Lim T. Mining "What they talk about" for a private healthcare service provider. *Archives of Business Research*. 2017; 5(5):1–22.
12. Nicholas, C. K. W., & Lee, A. S. H. (2017). Voice of customers: Text analysis of hotel customer reviews (cleanliness, overall environment & value for money). In *Proceedings of the 2017 International Conference on Big Data Research*, (2017), pp. 104-111. ACM. Crossref.
13. Paquette G. Visual knowledge modeling for semantic web technologies: Models and ontologies. Hershey, PA: Information Science Reference; 2010. Crossref 1, 2.
14. Raymond J. Mooney and Un Yong Nahm, "Text Mining with Information Extraction", *Multilingualism and Electronic Language Management: Proceedings of the 4th International MIDP Colloquium*, pp. 141-160, 2003. PMID:12543550
15. Schreiber G, Akkermans H, Anjewierden A, Hoog R, Shadbolt N, Velde V. *Knowledge engineering and management*. Cambridge, Mass: MIT Press; 2000.
16. Shilpa D. "A Review of Text Mining Techniques Associated with Various Application Areas" *International Journal of Science and Research (IJSR)*, 2015; 4. 2461-2466.
17. Sklar R. *Movie-made America: A cultural history of American Movies*. New York: Vintage Books; 1994. PMCID: PMC2568182. PMCID:PMC2568182
18. Talib R, Kashif M, Ayesha S, Fatima F. Text Mining: Techniques, applications and issues. *International Journal of Advanced Computer Science and Applications*. 2016; 7(11):1–5. Crossref.